74961_Auto_Edited.docx

Retrospective Study

Modified preoperative score to predict disease-free survival for hepatocellular carcinoma patients with surgical resections

Lai Y *et al*. Preoperative prediction of HCC survival.

Yin Lai, Jin-Chiao Lee, Hao-Chien Hung, Yu-Chao Wang, Chih-Hsien Cheng, Tsung-Han Wu, Chen-Fang Lee, Ting-Jung Wu, Hong-Shiue Chou, Kun-Ming Chan, Chen-Yao Kao, Wei-Chen Lee

Abstract

BACKGROUND

No prognostic models specific to hepatocellular carcinoma (HCC) receiving surgical resection have been considered strong and convincing enough for survival prediction thus far, let alone the models including only preoperative predictors. We derived a nomogram to predict disease-free survival in a previous study.

AIM

To simplify our score and compare research outcomes among other scoring systems.

METHODS

We retrospectively reviewed data from 1106 patients with HCC who underwent liver resection at the Linkou Chang Gung Memorial Hospital between April 2003 and December 2012. Multivariate analyses were conducted to identify the significant survival predictors. Homogeneity, Harrell's C-index, and Akaike information criterion were compared between our score, AJCC 8th edition, Tokyo score, and Taipei Integrated Scoring System (TTV-CTP-AFP model).

RESULTS

Among the 1106 patients, seven hundred and thirty-one (66.1%) had tumor recurrence at a median follow-up of 83.9 mo. Five risk factors were identified: Platelet count, albumin level, indocyanine green retention rate, multiplicity, and radiologic total tumor volume. Patients were divided into three risk groups, and the 5-year survival rates were 61.7%, 39%, and 25.7%, respectively. The C-index was 0.617, which was higher than the Tokyo score (0.613) and the Taipei Integrated Scoring System (0.562) and equal to the value of the AJCC 8th edition (0.617).

CONCLUSION

The modified score provides an easier method to predict survival. Appropriate treatment can be planned preoperatively by dividing patients into risk groups.

Key Words: Hepatocellular carcinoma; Preoperative; Prediction; Tumor recurrence

Lai Y, Lee JC, Hung HC, Wang YC, Cheng CH, Wu TH, Lee CF, Wu TJ, Chou HS, Chan KM, Kao CY, Lee WC. Modified preoperative score to predict disease-free survival for hepatocellular carcinoma patients with surgical resections. *World J Hepatol* 2022; In press

Core Tip: This retrospective study recruited over one thousand patients and developed a simple preoperative score to evaluate the recurrence risk of hepatocellular carcinoma after surgical resection. Despite the lack of pathological features, predictive power was satisfactory. Appropriate treatment can be planned preoperatively by dividing patients into risk groups.

INTRODUCTION

Hepatocellular carcinoma (HCC) is a complex malignant tumor. Associated with various clinical risk factors, HCC arises from a cirrhotic or non-cirrhotic liver with different degrees of viral or metabolic etiological exposure[1] and develops in molecular and intratumoral heterogeneities[2,3]. These reasons cause difficulty in developing staging systems for outcome prediction worldwide[4]. Although well-known conventional staging systems, such as Okuda[5], the AJCC 8th edition (TNM)[6], BCLC[7], JIS[8], and CLIP[9] are derived from large samples containing patients in early and advanced stages, all have limitations. So far, no prognostic models specific to HCC receiving surgical resection have been considered strong and convincing enough for survival prediction, let alone the models including only preoperative predictors.

During the past few decades, researchers have attempted to enhance the predictive power of models in five major ways. First, markers other than alpha fetoprotein (AFP) were identified that contribute to prognosis prediction, such as AFP-L3, glypican-3, cyclase-associated protein 2, and so forth[10]. Second, tumor size and numbers were replaced with total tumor volume (TTV), which is more representative of tumor burden presentation[11,12]. Third, models were developed for

specific groups of patients to increase prediction accuracy, such as hepatitis B virus (HBV)/ hepatitis C virus (HCV)-related[13,14], AFP-positive/negative[15], specific Child-Pugh classification, within/beyond the Milan criteria[13] and so on. Fourth, a more precise statistical method, such as a nomogram[16-18], has been prioritized. Finally, new risk factors have been sought; however, they proved difficult to identify.

Based on the above enhancement goals, we derived a preoperative nomogram to predict disease-free survival (DFS) using a multivariate Cox regression model[19]. Prognostic factors included viral hepatitis, platelet count, albumin, indocyanine green (ICG) retention rate, tumor multiplicity, and radiologic TTV. We chose AFP as the only tumor marker for survival prediction analysis because it is widely used and highly accessible compared to other enzymes, cytokines, or genetic biomarkers. However, an AFP cut-off value of 200 did not result in a satisfactory survival prediction. Finally, the patients were grouped into three categories: Low, intermediate, and high risk of recurrence. The high-risk group had a poor median DFS of 12.4 mo and with a 5-year DFS rate of only 21.1%. Despite the large number of subjects and very long-term follow-up in the former study, the lack of comparison with other staging systems limited its credibility. Thus, the aims of the present study were to collect data from a larger sample, simplify the score, and compare the research outcomes with those derived from other scoring systems.

MATERIALS AND METHODS

Study population and clinical characteristics

Patients with HCC who underwent surgical resection at the Linkou Chang Gung Memorial Hospital between April 2003 and December 2012 were recruited retrospectively. The diagnosis of HCC was pathologically confirmed. Laboratory data before primary liver resection (LR) were obtained from medical records. Preoperative computed tomography (CT)/ magnetic resonance imaging images were obtained for TTV calculation using the following formula: (length) × (width)^2 × 0.52, a modified method for ellipsoid volume measurement[20,21]. A total of 1106 subjects who had met the eligibility criteria were selected after excluding patients

with double malignancy, missing data, a positive pathological margin, or 30-d mortality like our previous study (Figure 1). The median follow-up was 83.9 mo. This study was approved by the local ethics committee of the Chang Gung Memorial Hospital.

Treatment and follow-up

LR was completed histologically when there was no evidence of distant metastasis. After surgery, the patients were followed up regularly by monitoring liver function tests, AFP levels, and liver ultrasonography every three months. Dynamic CT of the liver was performed if necessary. Tumor recurrence was defined using clinical, radiological, and/or pathological criteria similar to the initial HCC diagnosis. Disease-free survival was calculated based on the period between the date of surgery and tumor recurrence.

Statistical Analysis

Descriptive statistics for clinicopathological variables are presented. Statistical significance was defined as a P-value < 0.05. The optimal cutoff values of TTV were determined using the maximally selected rank statistics in R. The Kaplan-Meier method and log-rank test were used for DFS analysis. Significant variables associated with DFS in the univariate analysis were included in the multivariate Cox proportional hazards model. Scores were assigned to each prognostic predictor according to the results. The performances of the different scoring systems were compared using the likelihood ratio chi-square score for homogeneity, linear trend chi-square score, Harrell's concordance index for discriminatory ability, and Akaike information criterion (AIC) for prognostic stratification. All analyses were conducted using the SPSS software (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, version 20.0. Armonk, NY, United States: IBM Corp.) and R version 4.0.5 [R Core Team (2021)]. R: Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

RESULTS

Patient clinicopathologic characteristics

Demographic characteristics are shown in Table 1. The majority of the patients were men (78%) with viral hepatitis (83%). A higher percentage of patients were over the age of 55 (61%) and only 0.01% were Child C. A majority of patients had a lower international normalized ratio (INR) (91%), total bilirubin (91%), ICG clearance (68%), and higher albumin (92%) levels. Seventy-one percent of the patients had an AFP level < 200 ng/mL. Regarding tumor burden, more patients had solitary tumors (77%) and a radiologic TTV ≤ 32.0 cm³ (58%). Pathologically, fewer patients had liver cirrhosis (47%), tumor rupture (3%), Edmondson-Steiner grade III/IV (38%), or microvascular invasion (29%). A higher percentage of tumor capsules (83%) and pathological TTV ≤ 32.8 cm³ (59%) were noted. Seven hundred thirty-one (66.1%) patients had tumor recurrence at a median follow-up of 83.9 mo.

Risk factors identified in the preoperative prognostic model

After pooling data from the two databases, platelet count ($P = 0.003$), total bilirubin ($P = 0.032$), albumin ($P = 0.001$), ICG clearance rate ($P < 0.0001$), multiplicity of tumor ($P < 0.0001$), and radiologic TTV ($P < 0.0001$) were significantly associated with DFS in univariate analysis. Viral hepatitis, which was found to have predictive potential in a previous study, did not show prognostic significance in the univariate analysis ($P = 0.111$). Five predictors remained significant in multivariate analysis, including platelet count [$P = 0.001$, hazard ratio (HR) = 1.498, 95% confidence interval (CI):1.192-1.882], albumin ($P = 0.005$, HR = 1.462, 95% CI: 1.121-1.907), ICG clearance rate ($P = 0.001$, HR = 1.289, 95% CI: 1.104-1.507), multiplicity of tumor ($P < 0.0001$, HR = 1.694, 95% CI: 1.422-2.019), and radiologic TTV ($P < 0.0001$, HR = 1.743, 95% CI: 1.501-2.024) (Table 2). With these factors, the score was calculated by assigning 2 points for platelet count, multiplicity, and TTV and 1 point each for albumin and ICG according to the calculation of the regression coefficient formula. The percentages of patients with risk scores from 0 to 7 were 28.3%, 13.0%, 28.4%, 15.3%, 9.3%, 4.3%, 1.3%, and 0.1%, respectively. Patients with 0, 1-2, and 3-7 points were categorized into low-, intermediate-, and high-risk groups, according to the ascending possibility of the 16th, 50th, and 84th percentiles.

Radiological errors between CT and pathology

When radiological error of multiplicity was examined using a cross table, only one subject out of 1106 patients with solitary tumor was misdiagnosed with multiplicity on CT. In contrast, 51 subjects with multiple tumors were misdiagnosed with solitary tumors on CT. The diagnostic sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of CT were 79.9%, 99.9%, 99.5%, and 94.5%, respectively. The overall accuracy was 95.3%. As for optimal radiological TTV cut-off value (32.0 cm^3), the diagnostic sensitivity, specificity, PPV, and NPV of CT scan were 89.7%, 92.1%, 88.9%, and 92.7%, respectively, achieving accuracy of 91.1%.

Performance comparison of four prognostic models

The performance of our score was further compared with those of the AJCC 8th edition (TNM), Tokyo score, and Taipei Integrated Scoring System (TTV-CTP-AFP model). Figure 2 displays the survival curve of each group and the postoperative 1-, 3-, and 5-year DFS rates of the different scoring systems. There were statistically significant differences in long-term survival between the three groups. The 5-year DFS rates of our score from low-to high-risk groups were 61.7%, 39.0%, and 25.7%, respectively; AJCC 8th edition from stage IA to IIIB were 60.0%, 44.6%, 36.8%, 31.7%, and 21.2%, respectively; six groups of the Tokyo system were 76.0%, 48.3%, 45.0%, 29.9%, 18.9%, and 21.4%, respectively; and the five Taipei groups were 44.8%, 39.9%, 37.3%, 22.6%, and 40.0%, respectively. Table 4 illustrates the HR of the risk groups among the four scoring systems. The three groups of our score and the five groups of AJCC 8th edition appeared had growing risks according to HR. However, the highest risk groups in Tokyo and Taipei scores with lower HR (4.10 *vs* 4.14 in Tokyo; 1.26 *vs* 1.79 in Taipei) lost such a discrimination ability for risk stratification. Our score exhibited the highest likelihood ratio (χ2), linear trend (χ2), and lowest AIC value, indicating the best homogeneity, discriminatory ability, and prognostic prediction ability (Table 5). We also had an acceptable C-index value (0.617) equal to the AJCC 8th edition and superior to the Tokyo (0.613) and Taipei (0.562) scores.

DISCUSSION

Preoperative characteristic differences between two databases

In the nomogram of the preoperative prediction model that we modeled after the former database, TTV had the highest points of 100, and viral hepatitis was assigned 61 points. Viral hepatitis ranked fifth among only six risk factors above the ICG clearance level (39 points). Although the proportion of patients with or without viral hepatitis was similar between the two databases, this factor did not show a predictive potential in this study. In contrast, ICG remained significant and had the lowest regression coefficient, similar to our previous results. Notably, viral hepatitis remains the main cause of HCC in the Western Pacific Region, even with widespread HBV vaccination. However, the prevalence of viral hepatitis is relatively low in western countries. For example, only 3192 cases of acute hepatitis B and 4136 cases of acute hepatitis C were reported in the United States in 2019 (there are an estimated 257 million people living with HBV, and 71 million with HCV globally)[22]. In other words, without the factor of viral hepatitis, this score may be more applicable to western populations for DFS prediction. Additionally, a significantly lower percentage of multiplicity was observed in the current database. The annual number of cases of living-donor liver transplantation for HCC at our hospital has increased from 5 to approximately 30 over the past two decades. While the proportion of patients undergoing liver transplantation continues to rise, fewer patients with multiple tumors according to Milan criteria choose to receive LR. As for other preoperative variables, no patients had Child-Pugh class C in the newly collected data. More patients had better platelet counts, bilirubin, INR, albumin, and ICG clearance levels. Another popular predictor, neutrophil-to-lymphocyte ratio (NLR), was not included in the regression analysis in a previous study because of the large amount of missing data. The complete NLR compiled from the new database was not statistically significant in the univariate analysis (cut-off value: 2.5, P = 0.962).

Our score and the AJCC 8th edition were equally matched in predictive power but simpler

TTV and multiplicity ranked first and second, respectively, in the predictive power of our score. Because of our concern about possible radiology errors between CT scans and pathology, the probability was calculated. As indicated by our results, there was only a slight chance (0.49%) that multiple tumors would be mistaken for solitary tumors on CT. Approximately 18% of the patients were found to have multiple lesions when HCC was newly diagnosed. Eighty percent had identical pathological findings, but some daughter nodules that were difficult to detect on preoperative imaging caused diagnostic errors. Fifty-one subjects were missed out of 254 cases, with multiplicity confirmed by pathology. However, sensitivity (79.9%), specificity (99.9%), and overall accuracy (95.3%) remained highly satisfactory. Likewise, the CT scan performed remarkably well in distinguishing TTV. A possible reason for this finding is that the accuracy of CT scans was more limited in advanced HCC with a cirrhosis background. Patients who underwent LR in our hospital were mostly Child A, BCLC 0, or A without severe liver cirrhosis, leading to a more precise and accurate detection rate.

When comparing our score with the AJCC 8th edition, the low-risk group had a very close median DFS compared to the stage IA group, both exceeding 90%; the intermediate group had a similar median DFS of less than 40%, similar to the stage II group; and the high-risk group had a median DFS of less than 20% was between the stage IIIA and IIIB groups. In fact, for those who had recurrence in different groups, 28.2%, 52.5%, and 64.7% of patients had recurrence beyond the Milan criteria from the low-to high-risk groups, respectively. In this regard, patients with recurrence beyond the Milan criteria at variable stages of AJCC 8th edition with the following percentages were correlated with our risk groups: IA, 25.5%; IB, 41.8%; II, 61.4%; IIIA, 53.8%; and IIIB, 83.2%. Thus, the high-risk group of our score not only had an extremely high rate of recurrence of up to 79.4% but also had more advanced recurrence with limited treatment strategies. Simply put, even with a less delicate grouping, patients demanding adjuvant therapy and close monitoring could be accurately and conveniently selected from our score.

A better choice than the Tokyo or Taipei score for differentiating patient risk

The Tokyo scoring system, published by Shindoh *et al*[23] in 2020, uses three risk factors (tumor size > 2 cm, multiple lesions, and microvascular invasion) after pathological diagnosis. The score has the major advantage of simplicity over the classic prognostic staging systems, such as the TNM[6], Okuda[5], CLIP[9], JIS[8], CUPI[24], and GRETCH[25,26] but still requires pathological features. The Taipei Integrated System, developed by Yang-Ming University in 2010, was a true preoperative score derived from the Taiwanese population[27]. Although the Tokyo score had a C-index nearly comparable to our score, it was found to have an inferior discrimination ability and ambiguous hazard ratios in high-risk groups in this study, similar to the Taipei score.

The advantage of preoperative staging system in the near future

The age of multidisciplinary treatment is emerging, including targeted therapy, immunotherapy, and even cell therapy. Before reaching a consensus regarding adjuvant HCC therapy following resection, more evidence is needed. For instance, the STORM trial in 2016 noted that adjuvant sorafenib had no significant recurrence-free survival benefit[28] whereas a meta-analysis Huang *et al*[29] published in 2021 demonstrated that adjuvant sorafenib could not only prolong overall and recurrence-free survival but also reduce the recurrence rate. The effectiveness of adjuvant therapy, let alone the use of neoadjuvant therapy, remains controversial. Currently, neoadjuvant therapy has only been applied for disease downstaging to achieve potentially curative resection or tumor progression limitations to protect patients from exceeding transplant criteria[30]. Adjuvant therapy may be introduced as a neoadjuvant treatment to provide survival benefits or prevent recurrence. The preoperative staging system will play a vital role in risk stratification.

CONCLUSION

The modified preoperative score provides an easier way to predict disease-free survival for HCC patients with surgical resections. Despite the lack of pathological

features, predictive power was satisfactory. Appropriate preoperative treatment can be planned by simply dividing patients into three risk groups.

ARTICLE HIGHLIGHTS

Research background

No preoperative prognostic models specific to hepatocellular carcinoma patients surgical resection have been considered strong and convincing enough for survival prediction.

Research motivation

We previously derived a nomogram, but aimed simplify the score and compare it with other scoring systems.

Research objectives

To develop a simple preoperative score with satisfactory predictive power compared to postoperative scoring systems.

Research methods

Significant risk factors were identified using a multivariate Cox proportional hazards model. The homogeneity, Harrell's C-index, and Akaike information criterion of the different scoring systems were compared.

Research results

Five risk factors were identified and patients were divided into three risk groups. The C-index of our preoperative score was 0.617, which is equal to the value of the AJCC 8th edition.

Research conclusions

A modified score was established for survival prediction, and patients were divided into risk groups for preoperative treatment planning.

Research perspectives

Specific treatment or monitoring plan modifications for each risk group should be studied and potential correlation with survival benefit should be investigated.

74961_Auto_Edited.docx

ORIGINALITY REPORT

5%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.pubfacts.com Internet	41 words — **1%**
2	www.mdpi.com Internet	30 words — **1%**
3	journals.sagepub.com Internet	18 words — **1%**
4	www.science.gov Internet	14 words — **< 1%**
5	pubmed.ncbi.nlm.nih.gov Internet	13 words — **< 1%**
6	www.jove.com Internet	11 words — **< 1%**
7	www.researchgate.net Internet	11 words — **< 1%**
8	www.wjgnet.com Internet	11 words — **< 1%**
9	Adil Mohamed, Maya Shmulevitz. "p38 MAPK Signaling Enhances Reovirus Replication by Facilitating Efficient Virus Entry, Virus Capsid Uncoating and Post-Uncoating Steps", Cold Spring Harbor Laboratory, 2022	10 words — **< 1%**